Geophys. J. Int. (2000) **140,** 410–424

Three-dimensional magnetotelluric inversion using non-linear conjugate gradients

Gregory A. Newman[1] and David L. Alumbaugh[2]

[1] *Sandia National Laboratories, Geophysical Technology Department,* PO Box 5800, *Albuquerque,* NM 87185–0750, *USA.*
E-mail: ganewma@sandia.gov
[2] *University of Wisconsin-Madison, Department of Civil and Environmental Engineering,* 2258 *Engineering Hall,* 1451 *Engineering Drive, Madison,* WI 53706, *USA*

Accepted 1999 August 31. Received 1999 August 25; in original form 1998 November 2

SUMMARY

We have formulated a 3-D inverse solution for the magnetotelluric (MT) problem using the non-linear conjugate gradient method. Finite difference methods are used to compute predicted data efficiently and objective functional gradients. Only six forward modelling applications per frequency are typically required to produce the model update at each iteration. This efficiency is achieved by incorporating a simple line search procedure that calls for a sufficient reduction in the objective functional, instead of an exact determination of its minimum along a given descent direction. Additional efficiencies in the scheme are sought by incorporating preconditioning to accelerate solution convergence. Even with these efficiencies, the solution's realism and complexity are still limited by the speed and memory of serial processors. To overcome this barrier, the scheme has been implemented on a parallel computing platform where tens to thousands of processors operate on the problem simultaneously. The inversion scheme is tested by inverting data produced with a forward modelling code algorithmically different from that employed in the inversion algorithm. This check provides independent verification of the scheme since the two forward modelling algorithms are prone to different types of numerical error.

Key words: 3-D magnetotelluric modelling, inversion, non-linear conjugate gradients, preconditioning.

INTRODUCTION

An outstanding problem in the interpretation of 3-D magnetotelluric (MT) data sets has been the lack of robust and computationally efficient 3-D inversion schemes. Whilst 3-D forward modelling can be applied to these types of data sets, it is often too cumbersome to use for trial and error fitting of the observed data. Consequently, the interpretation of MT sounding data arising from 3-D geological settings has typically been based on 2-D approximations. Trial and error 2-D forward modelling fitting is sometimes used, but inversion methods are preferred because of their robustness, speed and ability to assess model uncertainties. A variety of 2-D inversion methods are available to the MT practitioner, including sharp-boundary and minimum-structure schemes (Smith *et al.* 1999; Rodi & Mackie 2000; Smith & Booker 1991; deGroot-Hedlin & Constable 1990; Smith 1988). In many instances accurate interpretation of 3-D geology can be achieved by the careful application of 2-D modelling and inversion (*cf.* Wannamaker *et al.* 1984). In other instances it may not, particularly in areas of highly complex 3-D geology.

The solution of the 3-D MT inverse problem is non-trivial because it is non-linear and very large in scale. It is also extremely ill posed due to the insufficient amounts of data that are typically collected. In an ill-posed problem non-physical-based models can be found that produce acceptable fits between the observations and predicted data. Fortunately, the ill-posed nature of the problem can be addressed using regularization techniques with constraints, but the large scale difficulty remains. Thousands of data points are required and thousands to tens of thousands of model parameters must be estimated in order to render an accurate 3-D map of the subsurface conductivity. In such an inverse problem it can be computationally prohibitive to form and solve repeatedly with direct methods a full least-squares system matrix to determine the model update. Because of these difficulties, Smith & Booker (1991) promoted a rapid relaxation technique for imaging 3-D MT data sets. The method approximates horizontal derivatives of the MT fields with their values calculated from the fields of the previous inversion iteration. This simplification uncouples the 3-D problem into a set of 1-D inverse problems, which can be quickly solved. Unfortunately, this simplification ignores

lateral effects in the computation of the data sensitivities, which can be significant in three dimensions, and can produce artefacts in the reconstructed conductivity.

Viable techniques to deal with the fully coupled inverse problem, which included full treatment of lateral 3-D effects, originated with the pioneering work of Mackie & Madden (1993) and Madden & Mackie (1989). Linear conjugate gradient (CG) methods were employed to minimize the data misfit by approximating it with series of convex quadratic models. By carefully implementing the CG algorithm, the difficulty of constructing, storing and solving a large and dense linear system of equations with expensive direct methods was avoided. The CG technique belongs to a class of iterative gradient methods that are computationally efficient for solving large-scale optimization problems because of their minimal storage requirements and the simplicity of their iteration.

Another critical strategy was the introduction of finite difference modelling techniques to simulate efficiently MT fields, sensitivities and functional gradients arising from complex 3-D geology. Over the last 10 years, 3-D MT finite difference modelling has matured and is now very well accepted within the MT community (*cf*. Smith 1996a; Weidelt 1995; Mackie *et al*. 1993; Madden & Mackie 1989).

The encouraging results of Mackie & Madden (1993) and Madden & Mackie (1989) have motivated us to consider non-linear optimization techniques for solving the 3-D MT inverse problem. We are interested in investigating such techniques to see if they are competitive with the linearized CG approach by sharing its key efficiencies and because they can be quite effective in dealing with the non-linear nature of the inverse problem. A survey of the non-linear optimization literature indicates that two iterative gradient methods, limited memory quasi-Newton and non-linear conjugate gradient (NLCG), appear to be promising candidates (Nocedal 1996; Gill *et al*. 1981). Here we will focus on the NLCG method because of its minimal storage requirements.

When Rodi & Mackie (2000) applied the NLCG method to the 2-D MT inverse problem, they obtained impressive results, including a significant reduction in the number of forward modelling applications. When the NLCG method is applied to the 3-D problem, we will show that only six forward modelling applications per frequency are usually required for the problem at a given iteration. Compared with competing formulations, we believe the scheme requires the fewest number of forward modelling applications per iteration. However, the question also rests on the total number of forward modelling applications needed over the entire inversion run and not just at a single iteration. If the NLCG method converges slowly, then the method's efficiency per iteration is irrelevant. Thus it is also important to find an effective preconditioner to speed up the convergence rate of the method in a global sense, as Rodi & Mackie (2000) demonstrated for the 2-D problem. Nevertheless, the complexity and realism of the images that can be produced by such a scheme is still limited by the speed and memory of serial processors. To overcome this barrier, parallel computers, where tens to thousands of processors operate on the problem simultaneously, can be employed, resulting in a significant reduction in execution times and a corresponding increase in model complexity.

In this paper, a parallel 3-D MT inverse solution will be formulated using preconditioned non-linear conjugate gradients. We will develop the solution by first discussing the solution of the forward problem, which will be used to compute predicted data and functional gradients. The mathematical formulation for solving the inverse problem will follow and will include a discussion on preconditioning. A critical test of the inversion scheme will then be presented by inverting data produced with a forward modelling code that differs substantially from that employed in the inversion algorithm. This check provides independent verification of the scheme since the two forward modelling algorithms are prone to different types of numerical error. We will also use this check to demonstrate the efficiency of preconditioning.

THE 3-D MAGNETOTELLURIC FORWARD PROBLEM

An important consideration in solving the 3-D MT inverse problem is that the forward modelling solution must be able to simulate correctly the responses arising from realistic 3-D geology. Parametrizations of hundreds of thousands of cells are typically required for these types of numerical simulations, and finite difference modelling techniques, as already mentioned, are very efficient for the task and will be employed here.

Assuming a harmonic time dependence of $e^{i\omega t}$, where $i=\sqrt{(-1)}$ and ω is the angular frequency, the electric field, $\mathbf{E}$, satisfies the vector equation

$$\nabla \times \mu_0/\mu \nabla \times \mathbf{E} + i\omega\mu_0\sigma\mathbf{E} = 0\,. \tag{1}$$

In this expression the electrical conductivity and magnetic permeability are denoted by σ and μ, respectively, and μ_0 is the magnetic permeability of free space. Dirichlet boundary conditions are applied to eq. (1), where the tangential electric field boundary values are specified. These boundary values arise from a plane wave, with a given source field polarization, propagating in layered or 2-D geological media assigned at the boundaries of the 3-D problem.

When eq. (1) is approximated with finite differences using a Yee (1966) staggered grid (Fig. 1) and symmetrically scaled (*cf*. Alumbaugh *et al*. 1996; Newman & Alumbaugh 1995), a linear system results:

$$\mathbf{KE} = \mathbf{S}\,. \tag{2}$$

The matrix $\mathbf{K}$ is complex-symmetric and sparse with 13 non-zero entries per row, and $\mathbf{S}$ is the source vector that depends on the boundary conditions and source field polarization. This system can be efficiently solved using the quasi-minimum residual (qmr) method, which belongs to the class of Krylov subspace techniques that are highly efficient in iteratively solving sparse linear systems. The reader is referred to Alumbaugh *et al*. (1996) and Newman & Alumbaugh (1995) for details on how this solution is implemented. Once the electric fields have been determined on the mesh, the magnetic fields, $\mathbf{H}$, can be determined from Faraday's law,

$$\mathbf{H} = \nabla \times \mathbf{E}/(-i\omega\mu)\,, \tag{3}$$

by numerically approximating the curl of the electric field at various nodal points. One can then interpolate either the electric or the magnetic field nodal values to the point of interest. Based on 3-D model comparisons with the integral

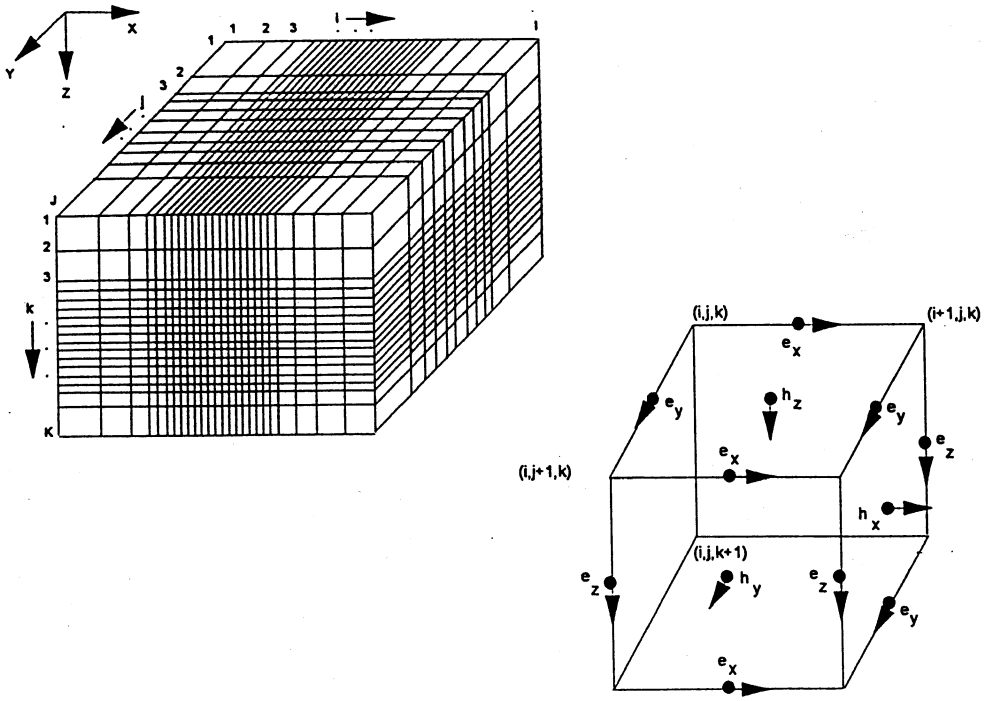

Figure 1. Mesh and staggered grid imposed upon the earth model to simulate 3-D MT fields. In the finite difference solution, the electric fields are sampled along cell edges and magnetic fields implicitly along the cell faces, where conductivity and permeability values are assigned to each cell.

equation solution of Xiong (1992), we estimate the error of the finite difference solution to be about 1 per cent in apparent resistivity and 1 degree in phase, for properly designed meshes.

Even with the benefits of a staggered grid, which implicitly enforces the divergence-free conditions on the fields, it is often necessary to explicitly enforce these conditions, where

$$\nabla \cdot \sigma \mathbf{E} = \mathbf{0} \tag{4}$$

and

$$\nabla \cdot \mathbf{E} = \mathbf{0} \tag{5}$$

in the Earth and air, respectively. This is accomplished in the numerical solution through a static divergence correction as frequencies approach the static limit. This correction, developed by Smith (1996b), can drastically reduce the time needed to solve eq. (2). It is implemented by adding a term to the electric field such that eqs (4) and (5) are periodically satisfied during the iterative solution of eq. (2).

THE 3-D MAGNETOTELLURIC INVERSE PROBLEM

Regularized least squares

All least-squares solutions begin by minimizing the difference between observed and predicted data, often subject to a constraint, which is employed to stabilize the inversion process. In the problem considered here, the inverse solution is constrained by Tikhonov regularization to remove its ill-posed nature (Tikhonov & Arsenin 1977). Regularization imposes an additional constraint on the model: reconstructions are required to be smoothed versions of the Earth's electrical conductivity, at the expense of an increase in the error (the difference between the measured and predicted data). We divide the Earth into M cells and assign to each cell an unknown conductivity value; the magnetic permeability, μ, is assumed to be constant within the Earth and is set to its free-space value from here on. Further, let $\mathbf{m}$ be a vector of length M that describes these values. We now form an objective functional, which combines the data error and model smoothness constraint in the following fashion:

$$\varphi = \sum_{n=1}^{2N} [(Z_n^{\text{obs}} - Z_n)/\varepsilon_n]^2 + \lambda \mathbf{m}^{\mathrm{T}} \mathbf{W}^{\mathrm{T}} \mathbf{W} \mathbf{m}, \tag{6}$$

where T denotes the transpose operator. In eq. (6), the terms that describe the observed and predicted data (magnetotelluric impedances), Z_n^{obs} and Z_n, are split into their real and imaginary parts. The summation over the first N data points is for the real parts of the impedance, and summation over the last N data points is for the imaginary parts. The impedance can be any particular component of the impedance tensor at any given measurement location, where the tensor is given by

$$\boldsymbol{Z} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix}. \tag{7}$$

Components of the tensor, which are complex, depend on the subsurface conductivity, and relate the measured horizontal electric and magnetic fields to each other. For example, $E_y = Z_{yx} H_x + Z_{yy} H_y$ gives the contribution to the y-component of the electric field arising from the x- and y-components of the magnetic field. Thus, for the horizontal components of the electric and magnetic field, $\mathbf{E}_{\mathrm{h}} = (E_x, E_y)^{\mathrm{T}}$ and $\mathbf{H}_{\mathrm{h}} = (H_x, H_y)^{\mathrm{T}}$, we have in general the relation $\mathbf{E}_{\mathrm{h}} = \boldsymbol{Z}\mathbf{H}_{\mathrm{h}}$. We also weight the data misfit in eq. (6) by the data error, ε_n, so that noisier data have less influence on φ.

The parameters that control model smoothness are the regularization matrix, **W**, which consists of a finite difference approximation to the Laplacian (∇^2) operator, and the regularization parameter λ, which is used to control the amount of smoothness to be incorporated into the model. Care needs to be taken in selecting this parameter. Large parameters will produce highly smooth models, but often these models do not fit the data. Small parameters give superior data fits but the resulting models can be too rough and non-physical. Our strategy is to run the inversion using several fixed values of λ and select the model that provides an acceptable match to the data within observational errors and yet yields the smoothest model. Because we plan to employ the non-linear conjugate gradient method to minimize the objective functional, λ should not be varied during the iteration procedure. To do so changes the definition of the objective functional as it is being minimized and would, in theory, invalidate the information determined from previous inversion iterations needed for computing the descent direction needed for the next model update. If λ were to be changed, it would be necessary to reinitialize the inversion procedure using the steepest descent direction, thereby discarding any information on prior descent directions, and use the model determined from the previous iteration as the new starting model. Note that these restrictions on λ are not necessary when linearized update methods are applied to the problem, where λ can be varied between updates (*cf.* deGroot-Hedlin & Constable 1990).

In small-scale inverse problems it may be feasible to determine the global minimum of eq. (6) with a brute force search in parameter space. For large-scale problems, as considered here, this is not an option. Instead, what is typically done is to set the gradient of the objective functional, $\nabla\varphi$, with respect to the model parameters to zero and find by some economical means those model parameters that satisfy a critical point. It is a well-known fact, however, that this critical point could be a local rather than a global minimum because we are solving a non-linear inverse problem. Moreover, we need to ensure that $\nabla\varphi = 0$ actually defines a minimum instead of a saddle point. This can be accomplished with second-derivative tests, which verify that the Hessian of the objective functional is symmetric positive definite at the critical point. Because the predicted impedances depend on the model **m** in a non-linear fashion, we will be required to solve $\nabla\varphi = 0$ using an iterative method. From our prior discussions, the non-linear conjugate gradient method is efficient for this task.

Non-linear conjugate gradients

The method of steepest descent is the easiest understood and simplest to implement of the gradient methods. Unfortunately it usually converges very slowly in practice. A better approach is the method of non-linear conjugate gradients, first proposed by Fletcher & Reeves (1964) for non-linear optimization, and later improved by Polyak & Ribière (1969). The method is closely related to the linear CG method of Hestenes & Stiefel (1952) and is identical if the objective functional is quadratic and exact line searches are made with the NLCG algorithm. Shown below is a flowchart of the Polyak & Ribière algorithm, which will be used in the analysis.

Algorithm flowchart

(1) set $i = 1$, choose initial model $\mathbf{m}_{(i)}$ and compute $\mathbf{r}_{(i)} = -\nabla\varphi(\mathbf{m}_{(i)})$
(2) set $\mathbf{u}_{(i)} = \mathbf{M}_{(i)}{}^{-1}\mathbf{r}_{(i)}$
(3) find $\alpha_{(i)}$ that minimizes $\varphi(\mathbf{m}_{(i)} + \alpha_{(i)}\mathbf{u}_{(i)})$
(4) set $\mathbf{m}_{(i+1)} = \mathbf{m}_{(i)} + \alpha_{(i)}\mathbf{u}_{(i)}$ and $\mathbf{r}_{(i+1)} = -\nabla\varphi(\mathbf{m}_{i+1})$
(5) stop when $|\mathbf{r}_{(i+1)}|$ is sufficiently small, otherwise go to step (6)
(6) set $\beta_{(i+1)} = (\mathbf{r}_{(i+1)}{}^{\mathrm{T}}\mathbf{M}_{(i+1)}{}^{-1}\mathbf{r}_{(i+1)} - \mathbf{r}_{(i+1)}{}^{\mathrm{T}}\mathbf{M}_{(i)}{}^{-1}\mathbf{r}_{(i)})/(\mathbf{r}_{(i)}{}^{\mathrm{T}}\mathbf{M}_{(i)}{}^{-1}\mathbf{r}_{(i)})$
(7) set $\mathbf{u}_{(i+1)} = \mathbf{M}_{(i+1)}{}^{-1}\mathbf{r}_{(i+1)} + \beta_{(i+1)}\mathbf{u}_{(i)}$
(8) set $i = i + 1$ and go to step (3)

For now we define $\mathbf{M}_{(i)}{}^{-1}$ and $\mathbf{M}_{(i+1)}{}^{-1}$ as identity matrices for all i, with units of $\mathrm{S}^2\,\mathrm{m}^{-2}$, given that the objective functional in eq. (6) is dimensionless. We will redefine these matrices later when preconditioning is discussed. To use the NLCG method sensibly requires that we carefully implement two calculations of the procedure efficiently. These are (1) calculate the gradient of the objective functional and (2) find the value of α that minimizes the expression $\varphi(\mathbf{m} + \alpha\mathbf{u})$ for specified model parameters **m** and a given conjugate search direction **u**.

Computation of the gradients

The gradient of the objective function in eq. (6) is formally written as

$$\nabla\varphi = \nabla\varphi_{\mathrm{d}} + \lambda\nabla\varphi_{\mathrm{m}}, \tag{8}$$

where φ_{d}, φ_{m} are functionals that relate to the data misfit and the model smoothness constraint, respectively. Evaluation of $\nabla\varphi_{\mathrm{m}}$ leads immediately to

$$\nabla\varphi_{\mathrm{m}} = 2\mathbf{W}^{\mathrm{T}}\mathbf{W}\mathbf{m}. \tag{9}$$

To compute a specific element of $\nabla\varphi_{\mathrm{d}}$, we redefine the observed and predicted impedances and their weighted difference in eq. (6) to be complex quantities, where

$$Z_n{}^{\mathrm{obs}} = Z_n{}^{\mathrm{obs}} + iZ_{n+N}{}^{\mathrm{obs}},$$
$$Z_n = Z_n + iZ_{n+N} \tag{10}$$

and

$$\Delta Z_n = (Z_n{}^{\mathrm{obs}} - Z_n)/\varepsilon_n{}^2 + i(Z_{n+N}{}^{\mathrm{obs}} - Z_{n+N})/\varepsilon_{n+N}{}^2.$$

Hence, for the kth model parameter we would have

$$\partial\varphi_{\mathrm{d}}/\partial m_k = -2\mathscr{R}e\sum_{n=1}^{N}(\Delta Z_n)^*\,\partial Z_n/\partial m_k, \tag{11}$$

where * indicates complex conjugation. In Appendix A, following Rodi (1976), sensitivity elements for the impedance tensor are shown to be

$$\begin{aligned}
\partial Z_{xxj}/\partial m_k &= -{}^1\mathbf{g}_j{}^{\mathrm{T}}{}_{xx}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_1) - {}^2\mathbf{g}_j{}^{\mathrm{T}}{}_{xx}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_2),\\
\partial Z_{xyj}/\partial m_k &= -{}^1\mathbf{g}_j{}^{\mathrm{T}}{}_{xy}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_1) - {}^2\mathbf{g}_j{}^{\mathrm{T}}{}_{xy}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_2),\\
\partial Z_{yxj}/\partial m_k &= -{}^1\mathbf{g}_j{}^{\mathrm{T}}{}_{yx}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_1) - {}^2\mathbf{g}_j{}^{\mathrm{T}}{}_{yx}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_2),\\
\partial Z_{yyj}/\partial m_k &= -{}^1\mathbf{g}_j{}^{\mathrm{T}}{}_{yy}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_1) - {}^2\mathbf{g}_j{}^{\mathrm{T}}{}_{yy}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_2).
\end{aligned} \tag{12}$$

These sensitivities quantify small changes in the tensor elements at location j due to small changes in the kth model parameter. $\mathbf{K}^{-1}$ denotes the inverse of the finite difference stiffness matrix

employed in the solution of the forward problem in eq. (2) and depends implicitly on frequency. The electric fields, $\mathbf{E}_1$ and $\mathbf{E}_2$, in eq. (12) are also determined from the forward problem; they provide the two source polarizations necessary to define the impedance tensor at each frequency. The eight vectors $({}^1\mathbf{g}_j{}^{\mathrm{T}}{}_{xx}, {}^2\mathbf{g}_j{}^{\mathrm{T}}{}_{xx}, {}^1\mathbf{g}_j{}^{\mathrm{T}}{}_{xy}, {}^2\mathbf{g}_j{}^{\mathrm{T}}{}_{xy}, {}^1\mathbf{g}_j{}^{\mathrm{T}}{}_{yx}, {}^2\mathbf{g}_j{}^{\mathrm{T}}{}_{yx}, {}^1\mathbf{g}_j{}^{\mathrm{T}}{}_{yy}, {}^2\mathbf{g}_j{}^{\mathrm{T}}{}_{yy})$ specified in Appendix A are linear combinations of vectors that interpolate electric and magnetic fields for the two source polarizations from the forward modelling finite difference grid to the receiver site at location j.

Combining eqs (11) and (12) we have

$$\partial\varphi_{\mathrm{d}}/\partial m_k = 2\mathscr{R}e\left[\sum_{n=1}^{N}(\Delta Z_n)^{*1}\mathbf{g}_n{}^{\mathrm{T}}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_1) + \sum_{n=1}^{N}(\Delta Z_n)^{*2}\mathbf{g}_n{}^{\mathrm{T}}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_2)\right], \quad (13)$$

where the vectors, ${}^1\mathbf{g}_n{}^{\mathrm{T}}$ and ${}^2\mathbf{g}_n{}^{\mathrm{T}}$ are determined by selecting from eq. (12) the component of the impedance that is being treated according to the summation index. For example, if we are treating Z_{xyj} as the nth data point, then ${}^1\mathbf{g}_n{}^{\mathrm{T}} = {}^1\mathbf{g}_j{}^{\mathrm{T}}{}_{xy}$ and ${}^2\mathbf{g}_n{}^{\mathrm{T}} = {}^2\mathbf{g}_j{}^{\mathrm{T}}{}_{xy}$.

It can be shown that the number of forward solutions needed to evaluate the gradient in eq. (13) is four per frequency. We define one forward problem at each distinct frequency and mode since eq. (2) is solved separately for these quantities. Thus, two forward solutions are required to obtain the electric fields $\mathbf{E}_1$ and $\mathbf{E}_2$ and two additional forward solutions are needed to compute the fields arising from the following source distributions at each frequency:

$${}^1\mathbf{g}^{\mathrm{T}} = -2\sum_{n-1}^{N}(\Delta Z_n)^{*1}\mathbf{g}_n{}^{\mathrm{T}} \quad (14)$$

and

$${}^2\mathbf{g}^{\mathrm{T}} = -2\sum_{n-1}^{N}(\Delta Z_n)^{*2}\mathbf{g}_n{}^{\mathrm{T}}. \quad (15)$$

Thus if we set

$${}^1\mathbf{v}^{\mathrm{T}} = {}^1\mathbf{g}^{\mathrm{T}}\mathbf{K}^{-1} \quad (16)$$

and

$${}^2\mathbf{v}^{\mathrm{T}} = {}^2\mathbf{g}^{\mathrm{T}}\mathbf{K}^{-1}, \quad (17)$$

we are then required to solve the two following forward problems:

$$\mathbf{K}\,{}^1\mathbf{v} = {}^1\mathbf{g} \quad (18)$$

and

$$\mathbf{K}\,{}^2\mathbf{v} = {}^2\mathbf{g}. \quad (19)$$

Here we utilized the fact that $\mathbf{K}^{\mathrm{T}} = \mathbf{K}$. In solving these forward problems we also employ the static divergence correction to accelerate their solutions at the lower frequencies. Here eq. (4) is modified to

$$\nabla\cdot\sigma\mathbf{E} + \nabla\cdot\mathbf{J} = \mathbf{0}, \quad (20)$$

where $\mathbf{J}$ is the source current density given by either ${}^1\mathbf{g}$ or ${}^2\mathbf{g}$. Additional details of the modified static divergence procedure can be found in Newman & Alumbaugh (1996) and Alumbaugh & Newman (1996).

We verified the accuracy of eq. (13) by performing a simple test proposed by Gill *et al.* (1981). This test is straightforward and computes the projected gradient using a finite difference approximation. Let $\mathbf{m}$ denote the point in model space where we evaluate the gradient. Choose a small scalar h and a random vector $\mathbf{p}$, of unit length, such that the elements of $\mathbf{p}$ are all similar in magnitude. It should then hold that the gradient projected along the direction $\mathbf{p}$ will approximately satisfy

$$(\nabla\varphi_{\mathrm{d}})^{\mathrm{T}}\mathbf{p} \approx [\varphi_{\mathrm{d}}(\mathbf{m}+h\mathbf{p}) - \varphi_{\mathrm{d}}(\mathbf{m})]/h. \quad (21)$$

A series of tests showed agreement to within 1 per cent when eq. (13) was substituted into the left-hand side of eq. (21).

The line search

A line search is needed to find α that will minimize $\varphi(\mathbf{m}+\alpha\mathbf{u})$. There are a number of strategies to carry out the line search, where the number of forward modelling applications varies. An economical approach, employed here and presented in detail in Appendix B, is to find α such that $\varphi(\mathbf{m}+\alpha\mathbf{u})$ is decreased sufficiently along the search direction $\mathbf{u}$. Quadratic curve fitting is also employed in an attempt to refine α so that an even greater reduction in φ is possible. Here, we use functional and derivative information at one point and functional information at another, and then fit a quadratic through these points to estimate the step to the minimum. Since objective functional and gradient information is already available at the descent point $\mathbf{m}$ from the prior NLCG iteration, we only need to evaluate φ at the second point along the direction of descent. Thus, the cost of the quadratic line search is two additional forward modelling applications per frequency. While it may seem preferable to use higher-order polynomial interpolation (e.g. cubic) to refine the line search, this can be expensive because it incurs additional functional and possible gradient evaluations (*cf.* Acton 1990). A good preconditioner can help to compensate for the approximate line search procedure employed here without recourse to additional forward modelling and is discussed below.

If the functional increases at the estimated minimum point, a backtracking strategy is invoked from this point until a sufficient decrease in the functional is observed. Backtracking can be done because $\mathbf{u}$ is a descent direction, since $\mathbf{u}\cdot\nabla\varphi < 0$ at $\varphi(\mathbf{m})$. We have found in practice that backtracking is typically not required in our scheme.

Preconditioning

The findings of Rodi & Mackie (2000) showed that convergence rates for the 2-D MT inverse problem using non-linear conjugate gradients can be significantly improved with preconditioning. Nevertheless, as Nocedal (1996) pointed out, the use of a preconditioner with a NLCG scheme can result in a conjugate search direction no longer being a descent direction and requires reinitializing using the steepest descent direction, which severely impairs the algorithm's efficiency. Whilst it is possible to take precautions to prevent this by ensuring that the preconditioner is positive definite, the use of an approximate line search procedure, as used in this paper, can still produce a search direction that is not a descent direction. In spite of these dangers, an effective and robust preconditioner is important, since it could significantly reduce solution run times for the 3-D problem.

The convergence of the NLCG method can be accelerated by choosing a preconditioner $\mathbf{M}_{(i)}$ that approximates the Hessian of the objective functional, where $\mathbf{M}_{(i)}{}^{-1}$ requires minimal storage and $\mathbf{M}_{(i)}{}^{-1}\mathbf{r}$ is easy to compute. $\mathbf{M}_{(i)}$ can be fixed or updated at each iteration of the procedure. We prefer to update $\mathbf{M}_{(i)}$, because the Hessian is not constant, except for points near the minimum of the objective functional. When the Hessian is constant the objective functional can be represented by a quadratic function in the model parameters.

Whilst it is too expensive to compute or approximate the full Hessian, it is possible to compute its diagonal for use as a preconditioner. This information is accessible from computations involved in the NLCG iteration. Using a quasi-Newton formula, known as the Broyden–Fletcher–Goldfarb–Shanno (BFGS) update (*cf.* Dennis & Schnabel 1996), we derive a recurrence relation for the diagonal of the Hessian even if its off-diagonal components are unknown. The diagonal is given by

$$\mathbf{M}_{(i+1)} = \mathbf{M}_{(i)} + \nabla\varphi(\mathbf{m}_{(i)})\nabla\varphi(\mathbf{m}_{(i)})^{\mathrm{T}}/\nabla\varphi(\mathbf{m}_{(i)})^{\mathrm{T}}\mathbf{u}_{(i)} + \mathbf{y}_{(i)}\mathbf{y}_{(i)}{}^{\mathrm{T}}/\alpha_{(i)}\mathbf{y}_{(i)}{}^{\mathrm{T}}\mathbf{u}_{(i)}, \quad (22)$$

with $\mathbf{y}_{(i)} = \nabla\varphi(\mathbf{m}_{(i+1)}) - \nabla\varphi(\mathbf{m}_{(i)})$. The use of this update as a diagonal preconditioner in the NLCG inversion scheme is justified because in the quadratic case with exact line searches identical search directions are generated by the BFGS and NLCG algorithms (Gill *et al.* 1981). Additional justification that eq. (22) should be an effective preconditioner comes from the optimization literature, where the Newton search direction, $\mathbf{s} = \mathbf{M}_{(i)}{}^{-1}\mathbf{r}$, if practical to compute, is the optimum choice for functional minimization. Since full evaluation of $\mathbf{M}_{(i)}{}^{-1}$ is not realistic, its evaluation using only diagonal entries is an obvious choice. Thus the effect of the preconditioner on the NLCG scheme is to attempt to make it behave more like Newton's method. When $\mathbf{M}_{(i+1)}$ is updated from eq. (22) it is important to ensure that it is positive definite so that the search direction determined from the preconditioned NLCG algorithm is a descent direction. Therefore, whenever any diagonal entry of $\mathbf{M}_{(i+1)}$ is negative, it is not updated. Note also that at the first iteration $\mathbf{M}_{(1)}$ is set to the identity matrix.

Logarithmic parameters

Inverting for logarithmic parameters is desirable since it restricts the parameters to be positive quantities, which is a physical requirement for the electrical conductivity. Reformulation of the inverse problem for logarithmic parameters with lower bounds requires that elements of $\mathbf{m}$ be redefined as $m'_k = \ln[(m_k - lb_k)/m_0]$ in eqs (6) and (9), where $m_k > lb_k$, with $lb_k > 0$ (Newman & Alumbaugh 1997) and $m_0 = 1\ \mathrm{S\,m^{-1}}$. The gradient in eq. (11) is also modified to read

$$\partial\varphi_{\mathrm{d}}/\partial m'_k = -2\mathscr{R}e\sum_{n=1}^{N}(m_k - lb_k)/m_0(\Delta Z_n)^*\,\partial Z_n/\partial m_k. \quad (23)$$

A corresponding modification also applies to eq. (13). Once m'_k is updated in the NLCG iteration, the parameter components that are of interest follow from the expression

$$m_k = m_0\,\mathrm{e}^{m'_k} + lb_k. \quad (24)$$

PARALLEL IMPLEMENTATION

In spite of the aforementioned efficiencies of the NLCG method in limiting the number of forward modelling applications at each iteration of the procedure, 3-D forward modelling continues to be a bottleneck in achieving acceptable run times. The expensive components of the procedure, functional gradients and objective functional evaluations, are traced to the time required to solve the 3-D forward problem. This is also compounded because multiple inversion runs are needed to determine the optimal trade-off parameter and sometimes the correct data weighting. Therefore, any reduction in forward modelling run times would have a corresponding impact in reducing solution times for the inversion. Another difficulty with forward modelling is that to model a full 3-D MT survey realistically over all frequencies and spatial locations may in some instances require a forward model parametrization of over a million cells, which is impractical on serial workstations.

To alleviate these difficulties we have implemented the inversion scheme to run on a parallel computing platform. These platforms allow tens to thousands of processors to operate on the problem simultaneously and significantly reduce the solution time and greatly increase the complexity and realism of the geological models. Such platforms allow an entire 3-D MT data set to be analysed concurrently, instead of partially. In using these machines a given number of processors are assigned to each coordinate direction (nx in x, ny in y and nz in z) of the model. The total number of processors employed is equal to $nx \times ny \times nz$. Since each processor only needs to make calculations for a subset of the forward and inverse problem grids, and because the processors are making their calculations in parallel, the solution time is reduced by a factor approximately equal to the number of processors employed.

Newman & Alumbaugh (1997) and Alumbaugh *et al.* (1996) showed how to implement inverse and forward problems on parallel platforms for controlled-source electromagnetic applications. Generalization to the MT problem is a straightforward extension and the interested reader is referred to these papers for the additional details. Currently the MT inversion scheme has been implemented on the 9000-processor Teraflop machine available at Sandia National Laboratories. Preliminary analysis of a large field data set shows the advantages of a parallel version of the scheme, where 28 224 model parameters were estimated using approximately 13 000 data points. A 42-fold increase in speed was observed when 512 Pentium II processors were applied to the problem compared to a single-processor IBM RS-6000 590 workstation (one inversion iteration required nearly 16 hr on the IBM). If 1000 processors are employed the speed-up approaches two orders of magnitude. Note that the processor speed on the IBM workstation is roughly comparable to eight Pentium II processors.

SYNTHETIC EXAMPLE

We have selected a synthetic example that will test the robustness of the inversion scheme in the presence of near-surface statics as well as demonstrate the benefits of preconditioning. This uses a data set produced by the integral equation modelling code of Xiong (1992) and also provides an independent check on the forward solution. Data simulated with an integral equation code are susceptible to different types of numerical errors from the finite difference technique, which is employed within the inverse solution.

Consider the 3-D model in the leftmost column of Fig. 2. The model represents a $0.1\ \mathrm{S\,m^{-1}}$ cube embedded in a $0.01\ \mathrm{S\,m^{-1}}$

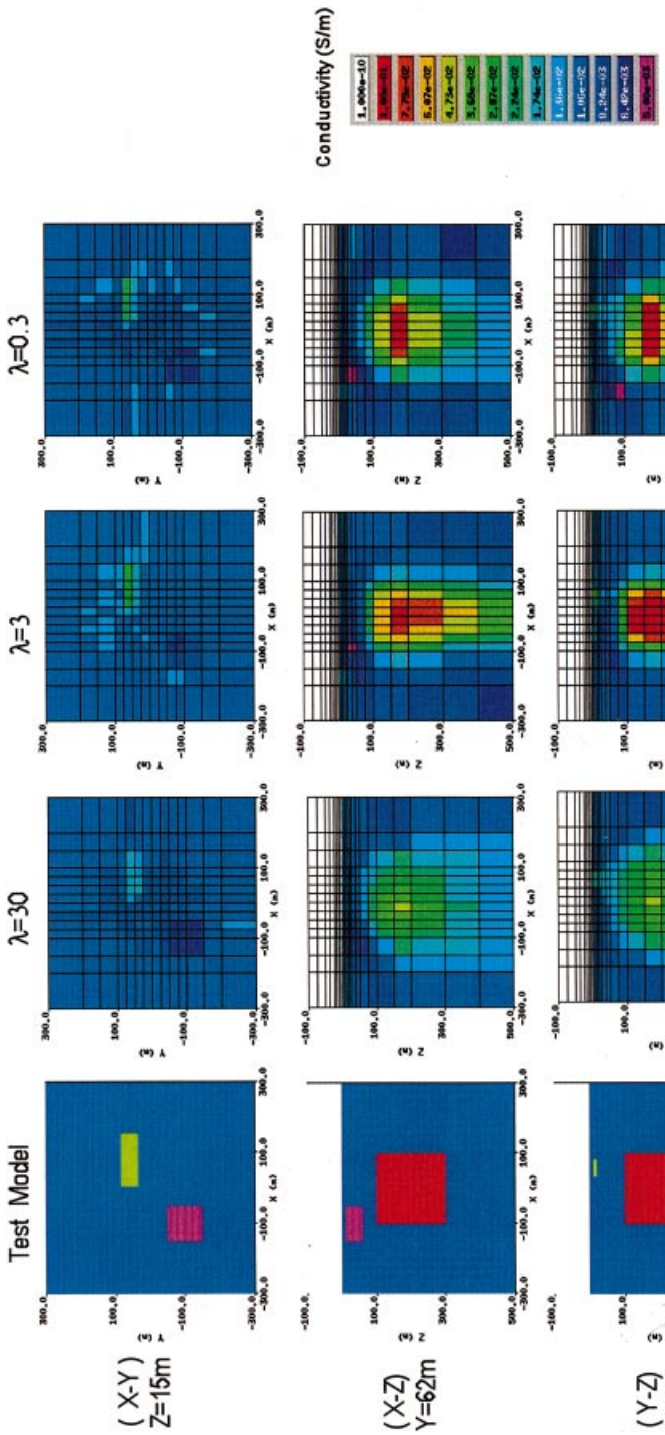

Figure 2. 3-D test model and corresponding inversion models obtained for three different trade-off parameters ($\lambda = 30$, 3 and 0.3) used in the test example. Different horizontal and vertical depth slices through the models are presented, where different projections (slices) of a given model are illustrated in column format. Along the top row are horizontal depth slices at 15 m for the different models. Two different vertical depth slices are shown in the second and third rows for $y = 62$ m and $x = 37$ m, respectively. Horizontal depth slices are also shown in the bottom row for a depth of 175 m. In the test model, the near-surface resistive body is violet, the conductive body is green and the deeper conductor is red. The mesh used in the inversion is also indicated and a conductivity colour bar scale is shown to the right.

half-space. The horizontal and vertical dimensions of the body are 200 m on a side. Its depth of burial is 100 m. Two near-surface bodies (one conductive and the other resistive) have also been included in the model to demonstrate the inversion code's capability to invert data that are affected by static shifts. Here the depth of burial and extent of the near-surface conductor are both 10 m. For the resistor they are 10 and 50 m, respectively. However, the finite difference mesh does not coincide exactly with the edges of the bodies in the model; offsets as large as 5 m occur between the mesh and the edges of the near-surface bodies. This supplies an additional realistic check on the inversion scheme as in the field it is generally impossible to align the mesh exactly with geological features.

121 MT soundings at 16 frequencies, ranging from 4000 to 4 Hz, were simulated using the integral equation solution of Xiong (1992); each sounding provides all four elements of the impedance tensor. The receivers are located at the Earth's surface ($z = 0$ m) and are uniformly spaced at 50 m intervals in both the x- and y-coordinate directions, ranging from -250 to $+250$ m. Thus the total number of data points (real and imaginary) used in the test is 15 488. Gaussian random noise was added to the data at the 1 per cent level. The data weights assumed in the inversion process are based on 1 per cent of the impedance of a 0.01 S m^{-1} half-space.

The mesh has been designed to increase with depth, approximately on a logarithmic scale, reflecting the loss of resolution with depth, which is an inherent limitation of the MT method. A fine mesh was also employed in the near surface because of the near-surface bodies. Overall 2548 and 28 224 cells were used to represent the respective inversion and forward modelling domains. The inversion domain is a rectangular block, 600 m in the horizontal directions and 500 m in the vertical direction, extending downwards from the Earth's surface. It is a subset of the forward modelling domain used to compute functional gradients and predicted data and is 1000 m on a side. The inversion started from a 0.01 S m^{-1} half-space and a lower bound constraint on the conductivity was set at 10^{-5} S m^{-1}. To invert the data set successfully required 6 hr of CPU time per inversion iteration on the IBM workstation. This compared with 30 min for 125 nodes on the Teraflop machine, which is more than an order of magnitude faster.

Three separate inversion runs with preconditioning were carried out for three different trade-off parameters of $\lambda = 30$, 3 and 0.3. The corresponding conductivity models are shown in Fig. 2. The inversions producing these images were terminated at different iterations in Fig. 3 to study how small error reductions at the later iterations affect the final model. For none of the three different regularization parameters (30, 3 and 0.3) were we able to reduce the objective functional and squared error (data misfit) to the assumed noise level in Fig. 3. This will be discussed in more detail below. We have also included the squared error plot in Fig. 3 to examine the quality of the fit between the observations and the predicted data. Because of the regularization term, the objective functional is larger than the squared error after the first inversion iteration and is a biased measure of the data misfit. On the other hand, it is the quantity that is actually being minimized by the inversion process. It is interesting to note that at the terminal stages of the inversion process, the objective functional is monotonic in the trade-off parameter; smaller trade-off parameters produce smaller values in the objective functional. However, this is not necessarily the case with the squared error.

The model in Fig. 2 for $\lambda = 30$ shows that we have recovered information on both near-surface bodies. Not shown, however, is the image illustrating that the near-surface bodies have been projected onto the Earth's surface. Because the highest-frequency component in the data set is only 4 kHz, this frequency is insufficient to resolve the true depths of the near-surface bodies, since the skin depth is 79 m in 0.01 S m^{-1} material. We have also recovered information on the deeper conductor's geometry (depth of burial and depth and lateral extents). The maximum conductivity estimates, however, are about a factor of two too small due to the size of the trade-off parameter imposed in the solution when compared with solutions that use smaller parameters below.

The models obtained with the smaller trade-off parameters in Fig. 2 are rougher and show more variability in the conductivity structure, as expected. Nevertheless, the models also show evidence of the near-surface structures. The presence of the deeper conductor is also clearly indicated, but the maximum conductivity estimates are a factor of two too large. The conductivity estimates indicated by the red coloured cells for these models range between 0.2 and 0.08 S m^{-1}. An interesting difference between these two models is the estimated depth to the base of the deeper conductor. The model along the Y–Z cross-section, with $\lambda = 3$, shows the conductor to extend to 500 m depth compared to 400 m depth when $\lambda = 0.3$. This difference can be explained with the number of NLCG iterations employed to construct the respective models. 50 iterations were used to obtain the model for $\lambda = 3$, whereas 27 iterations were used for the model with $\lambda = 0.3$. At the 20th iteration, when $\lambda = 3$, the base of the deeper conductor is better resolved, as is its conductivity (Fig. 4). This indicates that the benefit of a small reduction in the error level at the later iterations can be offset by extraneous structure in the final image. It could also indicate that the inversion process is actually fitting noise in the data, which is not indicated in Fig. 3, since the objective function and squared error lie above the desired target misfit.

The failure to achieve the target misfit in Fig. 3 for all three regularization parameters is of some concern. Newman & Alumbaugh (1996) found a similar result when inverting controlled-source crosswell electromagnetic data generated by the integral equation technique. They attributed the discrepancy to numerical differences between the forward modelling schemes employed to generate synthetic data and that employed within the inversion. Recall that we estimate the error between the two forward modelling approaches to be about 1 per cent. To test this hypothesis we repeated the inversion with 5 per cent noise added to the data for a regularization parameter of 3. Because the solution converged to only a slightly reduced objective functional value and squared error (φ_d), which were still above the target misfit, we do not believe that this entirely explains the observed phenomena. Another possible explanation is that we cannot properly solve for the static effects caused by the two near-surface bodies as the inversion grid is not exactly aligned with the boundaries of these targets. Finally, it may simply be that as we approach the minimum, the convergence becomes very slow, reminiscent of the steepest descent method, which the NLCG technique should in theory overcome. More probably it is a combination of all three of these factors. This is supported by a test where synthetic data were generated for the deeper body only, using the finite difference

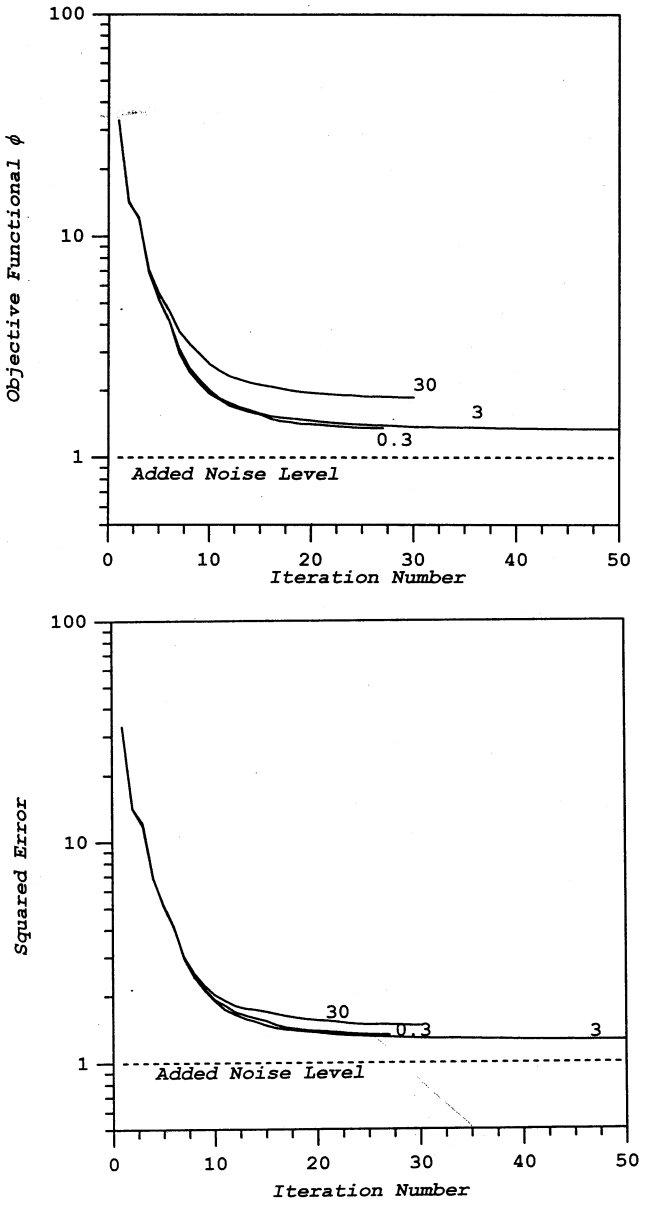

Figure 3. Plots of solution convergence (objective functional and squared error) for three different trade-off parameters of 30, 3 and 0.3. The difference between the objective functional and squared error plots is that the former includes the regularization term. As expected, a trade-off parameter of 0.3 produces the smallest realization of the objective functional at the terminal stages of the inversion process. In contrast, $\lambda = 3$ produces the smallest squared error measure.

scheme as employed in the inversion routine. Although convergence to the target noise level was achieved, it still took over 50 inversion iterations to obtain this result. Nevertheless the solution appears to be converging in Fig. 3, even though not to the desired target misfit. Moreover, a comparison between observed and predicted impedances, corresponding to the model in the middle column of Fig. 4, shows very good agreement in all four impedance components. This is illustrated in Fig. 5 at 40 Hz; similar agreement is also observed at other frequencies. We find the agreement between the diagonal tensor elements (Z_{xx} and Z_{yy}) especially encouraging given that these quantities are 3-D indicators, which vanish over 1- and 2-D geological structures.

In Fig. 6, we compare the convergence rate for the case $\lambda = 3$ with and without preconditioning. Preconditioning is found to be beneficial, as far fewer iterations are necessary to drive the objective functional and squared error to a given level at the later iterations, compared with a solution that does not use it. The final error level for the solution without preconditioning required 33 iterations This same error level was achieved in only 20 iterations when the preconditioned form of the solution was employed. The model obtained without preconditioning is shown in the rightmost column of Fig. 4 and should be compared directly with the model in the centre column. In this example, the model obtained through preconditioning resolves the deeper conductor somewhat better.

From these results, it is clear that the diagonal components of eq. (22) are an effective preconditioner for the test example. Curvature information built up by eq. (22) scales and steers the conjugate search direction to be a better prediction of the minimum. It is in the later stages of the inversion procedure that the benefits of preconditioning become obvious. This comes as no surprise, since eq. (22) will converge to the Hessian of the objective functional close to the minimum. Nevertheless, our findings on effective preconditioners for the 3-D inverse problem are preliminary. Rodi & Mackie (1998) have shown that for the 2-D problem a fixed preconditioner is very effective; we found this type of preconditioner ineffective for the 3-D scheme presented here. Thus, we believe additional research is needed on effective preconditioners for the 3-D problem.

DISCUSSION

It is worthwhile and enlightening to compare the number of forward modelling applications necessary to solve the 3-D MT inverse problem using linear and non-linear conjugate gradients. If a 3-D algorithm of the type discussed in this paper is used to compute the solutions to the forward problem, Mackie & Madden (1993) show that a linear CG scheme needs six forward modelling applications per frequency at the first relaxation step. Subsequent steps require four forward modelling applications per frequency. If the number of linear relaxation steps becomes excessive, we estimate greater than three, then an NLCG scheme should be very competitive. If on the other hand fewer steps are necessary to produce an accurate model update, the computational advantage would favour a linear CG implementation. Moreover, in the case of an objective functional exhibiting quadratic form, there would be no need to even consider a NLCG scheme. The key point is that the objective functional in eq. (6) can be non-quadratic and the linear method could require many model updates along with many relaxation steps at each update to achieve a minimum. In this event, a significant forward modelling expense would be incurred, making a preconditioned NLCG scheme a much more efficient choice. One could reasonably argue, however, that an implementation of the NLCG algorithm with approximate line search procedures, as employed in this paper, may obviate this advantage. It would take a direct comparison between the non-linear and linear algorithms over a series of test models to determine if and when this is the case.

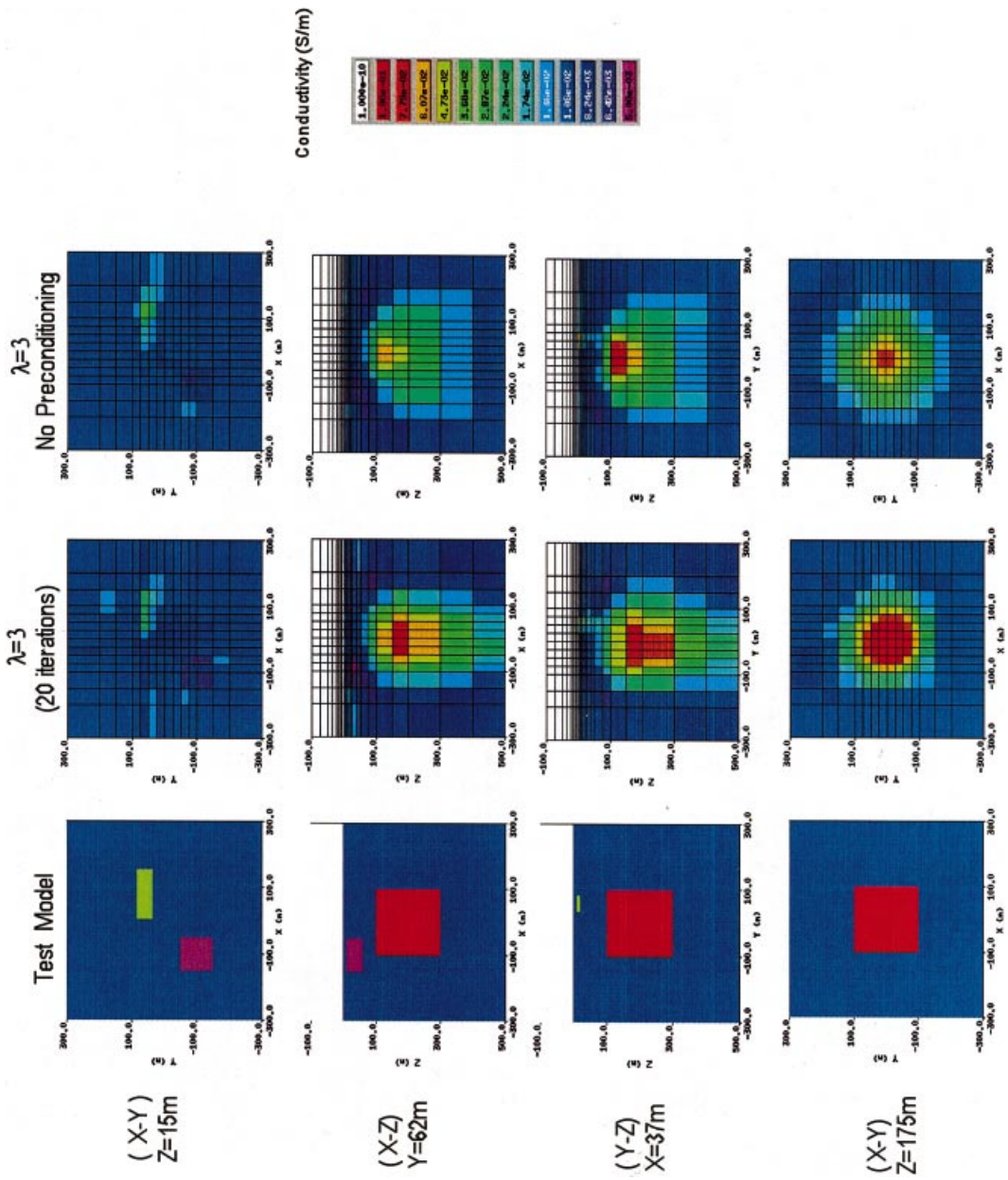

Figure 4. Different horizontal and vertical depth slices for the test model and inversion models in column format. Two inversion models are shown with and without preconditioning in the second and third columns. The centre or second column also presents image slices, where the inversion process has also been terminated at the 20th inversion iteration.

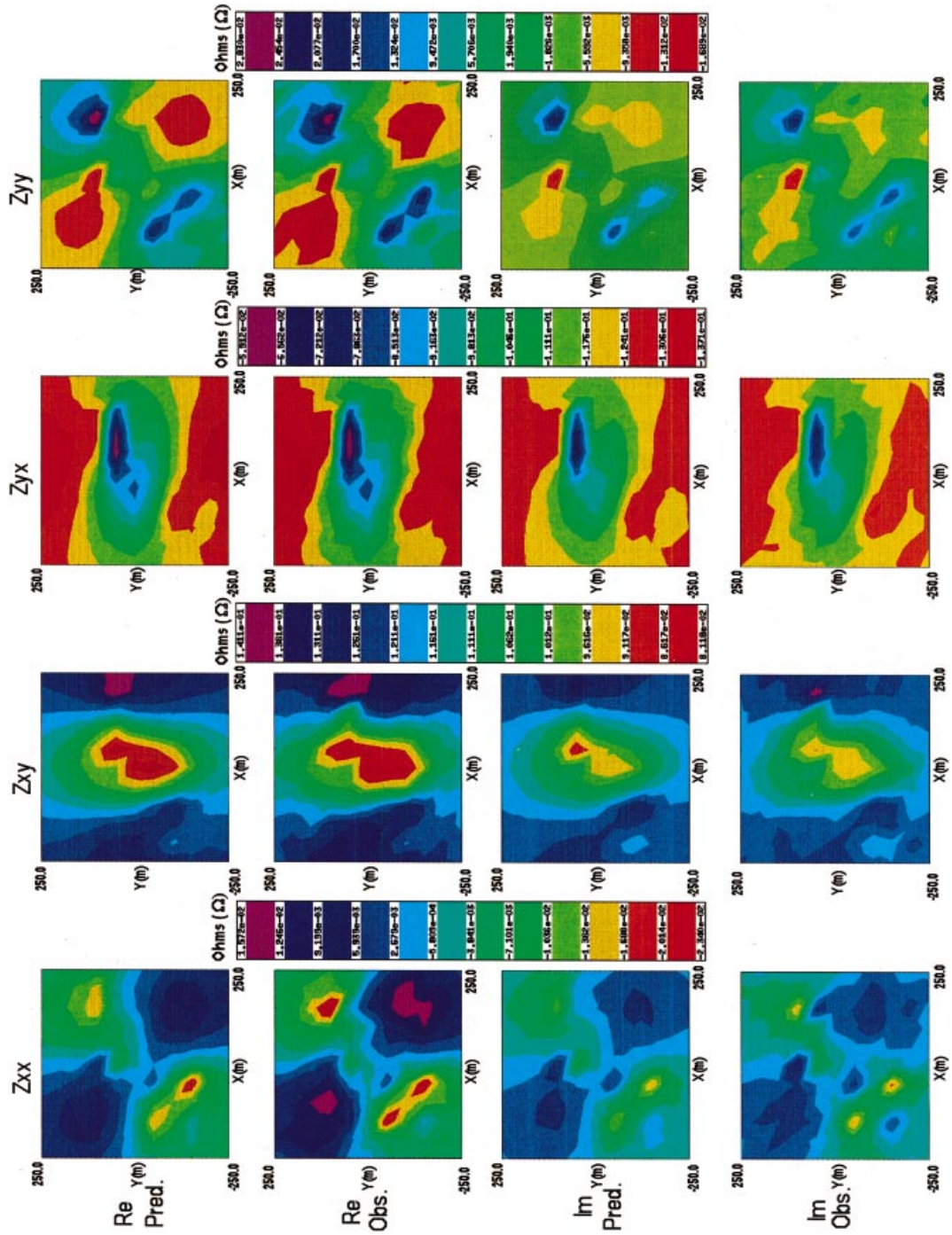

Figure 5. Calculated and observed impedance data (Z_{xx}, Z_{xy}, Z_{yx}, Z_{yy}) are illustrated at 40 Hz at the Earth's surface. Each column shows comparisons for the different impedance components along with its colour bar scale, immediately to the right. The top two rows in any given column show a comparison between the predicted (Pred.) and observed (Obs.) real (Re) impedance values and the bottom two rows show the corresponding results for the imaginary (Im) components.

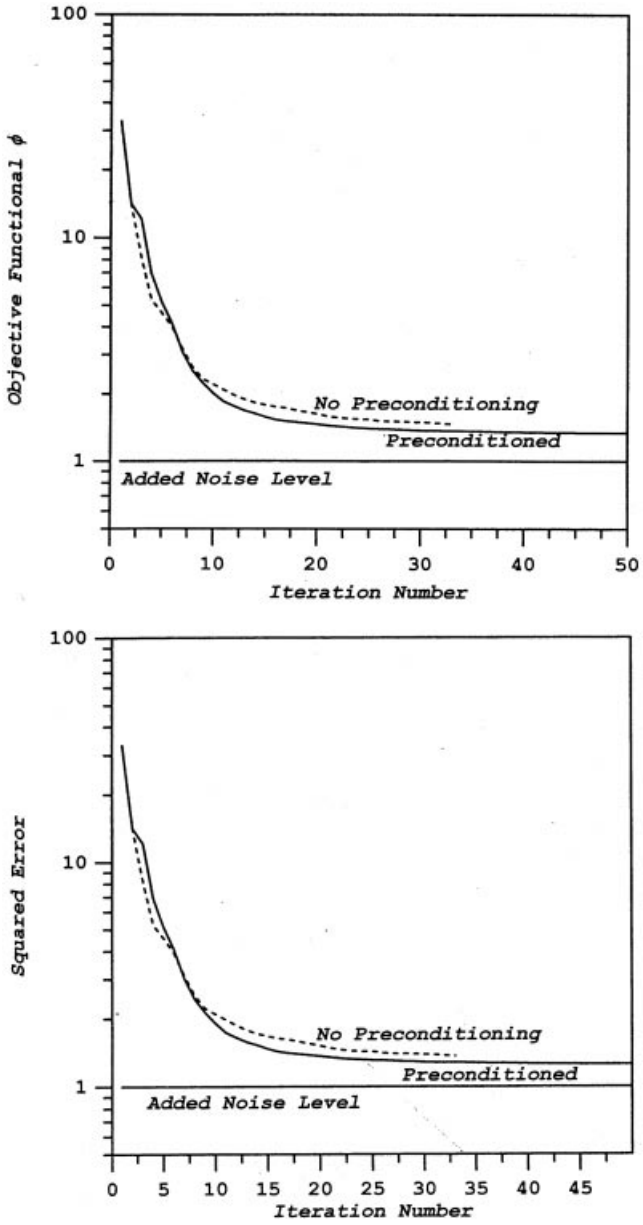

Figure 6. Plots of solution convergence for the objective functional and squared error with and without preconditioning for the test model in Fig. 2, where the trade-off parameter is set to 3.

CONCLUSIONS AND FUTURE WORK

In this paper we have presented a scheme to invert 3-D MT data sets on both serial and parallel computers. A preconditioned NLCG method has been employed because of its efficiency. It was demonstrated that only six forward modelling applications per frequency were typically needed to produce the model update. Application of the method on synthetic data shows that the method is viable.

To improve the convergence rate of the scheme, we are considering improvements in the line search procedures; this is in addition to improvements in preconditioning. The line search scheme employed in this paper calls only for a sufficient reduction in the objective functional, instead of determining its exact minimum as stipulated in the NLCG algorithm. A drawback to this type of line search is that a search direction can be produced that is no longer a descent direction (Gill *et al.* 1981) and requires restarting the algorithm with the steepest descent direction. A remedy for this situation, which is under current investigation, is to use a more elaborate line search procedure, at the expense of more functional evaluations, when it is determined that a quadratic model poorly approximates the objective functional.

In forthcoming papers we will apply the 3-D MT inversion scheme to field data sets. Exciting applications exist in marine magnetotellurics for oil prospecting (*cf.* Constable *et al.* 1998; Hoversten *et al.* 1998), mining and geothermal exploration, and deeper crustal and mantle investigations. We will also initiate a study on the validity of 2-D MT inversion over 3-D geological structures. This study will hopefully help to clarify when 3-D inversion is required.

ACKNOWLEDGMENTS

We express our gratitude to David Aldridge, Randy Macki, Bill Rodi and Martyn Unsworth for their critical reviews, which improved the paper. This work was performed at Sandia National Laboratories with funding provided by the United States Department of Energy's Office of Basic Energy Sciences, Division of Engineering and Geoscience. Sandia is a multiprogram laboratory operated by the Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04–94AL85000.

REFERENCES

Acton, F.S., 1990. *Numerical Methods that Work,* The Mathematical Association of America, Washington.

Alumbaugh, D.L. & Newman, G.A., 1996. Electromagnetic modeling of perfect conductors in an arbitrary host, *Extended Abstr. Soc. expl. Geophys. Ann. Mtng,* 978–981, Denver, CO.

Alumbaugh, D.L., Newman, G.A., Prevost, L. & Shadid, J.N., 1996. Three-dimensional wide and electromagnetic modeling on massively parallel computers, *Radio Sci.,* **31,** 1–23.

Constable, S.C., Orange, A., Hoversten, G.M. & Morrison, H.F., 1998. Marine magnetotellurics for petroleum exploration, Part I: a sea-floor equipment system, *Geophysics,* **63,** 816–825.

deGroot-Hedlin, C. & Constable, S.C., 1990. Occams's inversion to generate smooth, two-dimensional models from magnetotelluric data, *Geophysics,* **55,** 1613–1624.

Dennis, J.E., Jr & Schnabel, R.B., 1996. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,* Soc. Industrial appl. Math., Philadelphia.

Fletcher, R. & Reeves, C.M., 1964. Function minimization by conjugate gradients, *Comp. J.,* **7,** 149–154.

Gill, P.E., Murray, W. & Wright, M.H., 1981. *Practical Optimization,* Academic Press, London.

Hestenes, M.R. & Stiefel, E., 1952. Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bureau,* **49,** 409–436.

Hoversten, M.H., Morrison, H.F. & Constable, S.C., 1998. Part II: numerical analysis of subsalt resolution, *Geophysics,* **63,** 826–840.

Mackie, R.L. & Madden, T.R., 1993. Three-dimensional magnetotelluric inversion using conjugate gradients, *Geophys. J. Int.,* **115,** 215–229.

Mackie, R.L., Madden, T.R. & Wannamaker, P.E., 1993. Three-dimensional magnetotelluric modeling using difference equations—theory and comparisons to integral equation solutions, *Geophysics,* **58,** 215–226.

Madden, T.R. & Mackie, R.L., 1989. Three-dimensional magnetotelluric modeling and inversion, *Proc. IEEE,* **77,** 318–333.
Newman, G.A. & Alumbaugh, D.L., 1995. Frequency-domain modeling of airborne electromagnetic responses using staggered finite differences, *Geophys. Prospec.,* **43,** 1021–1042.
Newman, G.A. & Alumbaugh, D.L., 1996. Electromagnetic modeling of subsurface 3D structures, *Proc. Int. Geosci. Rem. Sens. Symp. (IGARSS),* 1941–44, Omaha, NB.
Newman, G.A. & Alumbaugh, D.L., 1997. Three-dimensional massively parallel electromagnetic inversion—I. Theory, *Geophys. J. Int.,* **128,** 345–354.
Nocedal, J., 1996. Conjugate gradient methods and nonlinear optimization, in *Linear and Nonlinear Conjugate Related Methods,* pp. 9–23, eds Adams, L. & Nazareth, J.L., Soc. Industrial appl. Math., Philadelphia.
Polyak, E. & Ribière, G., 1969. Note sur la convergence des méthods conjugées, *Rev. Fr. Inr. Rech. Oper.,* **16,** 35–43.
Rodi, W., 1976. A technique for improving the accuracy of finite element solutions for magnetotelluric data, *Geophys. J. R. astr. Soc.,* **44,** 483–506.
Rodi, W. & Mackie, R.L., 2000. Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion, *Geophysics,* in press.
Smith, J.T., 1988. Rapid inversion of multi-dimensional magnetotelluric data, *PhD thesis,* University of Washington, Seattle.
Smith, J.T., 1996a. Conservative modeling of 3-D electromagnetic fields; Part I: properties and error analysis, *Geophysics,* **61,** 1308–1318.
Smith, J.T., 1996b. Conservative modeling of 3-D electromagnetic fields; Part II: biconjugate gradient solution and an accelerator, *Geophysics,* **61,** 1319–1324.
Smith, J.T. & Booker, J.R., 1991. Rapid inversion of two- and three-dimensional magnetotelluric data, *J. geophys. Res.,* **96,** 3905–3922.
Smith, J.T., Hoversten, G.M., Gasperikova, E. & Morrison, H.F., 1999. Sharp boundary inversion of two-dimensional magnetotelluric data, *Geophys. Prospect.,* **47,** 469–486.
Tikhonov, A.N. & Arsenin, V.Y., 1977. *Solutions to Ill-Posed Problems,* John Wiley, New York.
Wannamaker, P.E., Ward, S.H. & Hohmann, G.W., 1984. Magnetotelluric responses of three-dimensional bodies in layered earths, *Geophysics,* **49,** 1517–1533.
Weidelt, P., 1995. Three-dimensional conductivity models: implications of electrical anisotropy, *Proc. Int. Sym. 3-D Electromagnetics,* 1–12, Schlumberger-Doll Research, Ridgefield, CT.
Xiong, Z., 1992. Electromagnetic modeling of 3-D structures by the method of system iteration using integral equations, *Geophysics,* **57,** 1556–1561.
Yee, K.S., 1966. Numerical solution of initial boundary problems involving Maxwell's equations in isotropic media, *IEEE Transactions on Antennas and Propagation,* **AP-14,** 302–309.

APPENDIX A: SENSITIVITY ELEMENTS FOR THE MT IMPEDANCE TENSOR

In order to derive the sensitivities, we need the predicted impedance tensor, where two orthogonal source polarizations are required. Following Mackie & Madden (1993), let the electric and magnetic fields for the two polarizations be denoted by $E_{x1}, E_{y1}, H_{x1}, H_{y1}$ and $E_{x2}, E_{y2}, H_{x2}, H_{y2}$. Thus

$$Z_{xx}=\frac{E_{x1}H_{y2}-E_{x2}H_{y1}}{H_{x1}H_{y2}-H_{x2}H_{y1}},\qquad Z_{xy}=\frac{E_{x2}H_{x1}-E_{x1}H_{x2}}{H_{x1}H_{y2}-H_{x2}H_{y1}},$$
$$Z_{yx}=\frac{E_{y1}H_{y2}-E_{y2}H_{y1}}{H_{x1}H_{y2}-H_{x2}H_{y1}},\qquad Z_{yy}=\frac{E_{y2}H_{x1}-E_{y1}H_{x2}}{H_{x1}H_{y2}-H_{x2}H_{y1}}. \tag{A1}$$

Differentiating each tensor element in eq. (A1) with respect to the kth model parameter, m_k, yields

$$\begin{aligned}
\partial Z_{xx}/\partial m_k &= [(H_{x1}H_{y2}-H_{x2}H_{y1})(H_{y2}\partial E_{x1}/\partial m_k+E_{x1}\partial H_{y2}/\partial m_k\\
&\quad -H_{y1}\partial E_{x2}/\partial m_k-E_{x2}\partial H_{y1}/\partial m_k)-(E_{x1}H_{y2}-E_{x2}H_{y1})\\
&\quad \times(H_{y2}\partial H_{x1}/\partial m_k+H_{x1}\partial H_{y2}/\partial m_k-H_{y1}\partial H_{x2}/\partial m_k\\
&\quad -H_{x2}\partial H_{y1}/\partial m_k)]/(H_{x1}H_{y2}-H_{x2}H_{y1})^2,\\
\partial Z_{xy}/\partial m_k &= [(H_{x1}H_{y2}-H_{x2}H_{y1})(H_{x1}\partial E_{x2}/\partial m_k+E_{x2}\partial H_{x1}/\partial m_k\\
&\quad -H_{x2}\partial E_{x1}/\partial m_k-E_{x1}\partial H_{x2}/\partial m_k)-(E_{x2}H_{x1}-E_{x1}H_{x2})\\
&\quad \times(H_{y2}\partial H_{x1}/\partial m_k+H_{x1}\partial H_{y2}/\partial m_k-H_{y1}\partial H_{x2}/\partial m_k\\
&\quad -H_{x2}\partial H_{y1}/\partial m_k)]/(H_{x1}H_{y2}-H_{x2}H_{y1})^2,\\
\partial Z_{yx}/\partial m_k &= [(H_{x1}H_{y2}-H_{x2}H_{y1})(H_{y2}\partial E_{y1}/\partial m_k+E_{y1}\partial H_{y2}/\partial m_k\\
&\quad -H_{y1}\partial E_{y2}/\partial m_k-E_{y2}\partial H_{y1}/\partial m_k)-(E_{y1}H_{y2}-E_{y2}H_{y1})\\
&\quad \times(H_{y2}\partial H_{x1}/\partial m_k+H_{x1}\partial H_{y2}/\partial m_k-H_{y1}\partial H_{x2}/\partial m_k\\
&\quad -H_{x2}\partial H_{y1}/\partial m_k)]/(H_{x1}H_{y2}-H_{x2}H_{y1})^2,\\
\partial Z_{yy}/\partial m_k &= [(H_{x1}H_{y2}-H_{x2}H_{y1})(H_{x1}\partial E_{y2}/\partial m_k+E_{y2}\partial H_{x1}/\partial m_k\\
&\quad -H_{x2}\partial E_{y1}/\partial m_k-E_{y1}\partial H_{x2}/\partial m_k)-(E_{y2}H_{x1}-E_{y1}H_{x2})\\
&\quad \times(H_{y2}\partial H_{x1}/\partial m_k+H_{x1}\partial H_{y2}/\partial m_k-H_{y1}\partial H_{x2}/\partial m_k\\
&\quad -H_{x2}\partial H_{y1}/\partial m_k)]/(H_{x1}H_{y2}-H_{x2}H_{y1})^2.
\end{aligned} \tag{A2}$$

An easy check is available on these expressions for a 2-D earth model. For both field polarizations we would have

$$(E_{x1}, H_{y1})\neq 0,\qquad (E_{y1}, H_{x1})=0$$

and

$$(E_{y2}, H_{x2})\neq 0,\qquad (E_{x2}, H_{y2})=0. \tag{A3}$$

Thus eqs (A2) reduce to

$$\begin{aligned}
&\partial Z_{xx}/\partial m_k=0,\\
&\partial Z_{xy}/\partial m_k=\{H_{y1}\partial E_{x1}/\partial m_k-E_{x1}\partial H_{y1}/\partial m_k\}/{H_{y1}}^2,\\
&\partial Z_{yx}/\partial m_k=\{H_{x2}\partial E_{y2}/\partial m_k-E_{y2}\partial H_{x2}/\partial m_k\}/{H_{x2}}^2,\\
&\partial Z_{yy}/\partial m_k=0,
\end{aligned} \tag{A4}$$

which are identical to expressions derived directly for the 2-D MT problem.

The sensitivities in eq. (A2) involve partial derivatives of the electric and magnetic fields. Following the ideas of Rodi (1976) we can relate these derivatives to the forward problem. Consider as an example the x-component of the magnetic field at location j for a given source polarization, which can be represented from Newman & Alumbaugh (1997) as

$$H_{xj}={}^{\mathbf{h}}\mathbf{g}^{\mathbf{T}}{}_{j(x)}\mathbf{E}. \tag{A5}$$

In this expression $\mathbf{E}$ is an electric field vector arising from a 3-D earth model for a specific source polarization and has dimensions of $N\mathrm{T}\times 1$, where $N\mathrm{T}$ represents the number of electric field unknowns that are determined from the finite

difference forward solution. The vector ${}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)}$ is an interpolator vector for the x-component of the magnetic field at the jth measurement point and is of dimension $1 \times N\mathrm{T}$ (T here denotes the transpose operator). This vector will interpolate the sampled fields on the forward modelling grid to the measurement point and numerically includes a curl operator that is applied to the electric field. With this definition an element of the Jacobian matrix is written for the x-component of magnetic field as

$$\partial H_{xj}/\partial m_k = {}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)}\partial\mathbf{E}/\partial m_k . \tag{A6}$$

From the forward problem, the electric fields are determined from the linear system given by eq. (2). Thus differentiating eq. (2) with respect to m_k yields,

$$\partial\mathbf{E}/\partial m_k = -\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}), \tag{A7}$$

and an element of the Jacobian matrix for the x-component of the magnetic field is written as

$$\partial H_{xj}/\partial m_k = -{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}). \tag{A8}$$

Similar expressions can be derived for the other electric and magnetic field components.

Using the above results, terms involving the electric and magnetic field partial derivatives can be substituted for in eq. (A2) to yield

$$\begin{aligned}
\partial Z_{xxj}/\partial m_k &= -{}^{1}\mathbf{g}_j{}^{\mathrm{T}}{}_{xx}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_1) - {}^{2}\mathbf{g}_j{}^{\mathrm{T}}{}_{xx}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_2),\\
\partial Z_{xyj}/\partial m_k &= -{}^{1}\mathbf{g}_j{}^{\mathrm{T}}{}_{xy}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_1) - {}^{2}\mathbf{g}_j{}^{\mathrm{T}}{}_{xy}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_2),\\
\partial Z_{yxj}/\partial m_k &= -{}^{1}\mathbf{g}_j{}^{\mathrm{T}}{}_{yx}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_1) - {}^{2}\mathbf{g}_j{}^{\mathrm{T}}{}_{yx}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_2),\\
\partial Z_{yyj}/\partial m_k &= -{}^{1}\mathbf{g}_j{}^{\mathrm{T}}{}_{yy}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_1) - {}^{2}\mathbf{g}_j{}^{\mathrm{T}}{}_{yy}\mathbf{K}^{-1}(\partial\mathbf{K}/\partial m_k\mathbf{E}_2),
\end{aligned} \tag{A9}$$

where $\mathbf{E}_1$ and $\mathbf{E}_2$ are the two electric field polarizations needed to define the impedance tensor and

$$\begin{aligned}
{}^{1}\mathbf{g}_j{}^{\mathrm{T}}{}_{xx} &= [(H_{x1}H_{y2} - H_{x2}H_{y1})(-H_{y2}{}^{\mathbf{e}}\mathbf{g}^{\mathrm{T}}{}_{j(x)} + E_{x2}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(y)})\\
&\quad + (E_{x1}H_{y2} - E_{x2}H_{y1})(-H_{x2}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(y)} + H_{y2}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)})]\\
&\quad /(H_{x1}H_{y2} - H_{x2}H_{y1})^2,\\
{}^{2}\mathbf{g}_j{}^{\mathrm{T}}{}_{xx} &= [(H_{x1}H_{y2} - H_{x2}H_{y1})(-E_{x1}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(y)} + H_{y1}{}^{\mathbf{e}}\mathbf{g}^{\mathrm{T}}{}_{j(x)})\\
&\quad + (E_{x1}H_{y2} - E_{x2}H_{y1})(-H_{y1}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)} + H_{x1}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(y)})]\\
&\quad /(H_{x1}H_{y2} - H_{x2}H_{y1})^2,\\
{}^{1}\mathbf{g}_j{}^{\mathrm{T}}{}_{xy} &= [(H_{x1}H_{y2} - H_{x2}H_{y1})(-E_{x2}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)} + H_{x2}{}^{\mathbf{e}}\mathbf{g}^{\mathrm{T}}{}_{j(x)})\\
&\quad + (E_{x2}H_{x1} - E_{x1}H_{x2})(-H_{x2}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(y)} + H_{y2}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)})]\\
&\quad /(H_{x1}H_{y2} - H_{x2}H_{y1})^2,\\
{}^{2}\mathbf{g}_j{}^{\mathrm{T}}{}_{xy} &= [(H_{x1}H_{y2} - H_{x2}H_{y1})(-H_{x1}{}^{\mathbf{e}}\mathbf{g}^{\mathrm{T}}{}_{j(x)} + E_{x1}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)})\\
&\quad + (E_{x2}H_{x1} - E_{x1}H_{y2})(-H_{y1}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)} + H_{x1}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(y)})]\\
&\quad /(H_{x1}H_{y2} - H_{x2}H_{y1})^2,\\
{}^{1}\mathbf{g}_j{}^{\mathrm{T}}{}_{yx} &= [(H_{x1}H_{y2} - H_{x2}H_{y1})(-H_{y2}{}^{\mathbf{e}}\mathbf{g}^{\mathrm{T}}{}_{j(y)} + E_{y2}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(y)})\\
&\quad + (E_{y1}H_{y2} - E_{y2}H_{y1})(-H_{x2}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(y)} + H_{y2}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)})]\\
&\quad /(H_{x1}H_{y2} - H_{x2}H_{y1})^2,\\
{}^{2}\mathbf{g}_j{}^{\mathrm{T}}{}_{xx} &= [(H_{x1}H_{y2} - H_{x2}H_{y1})(-E_{y1}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)} + H_{y1}{}^{\mathbf{e}}\mathbf{g}^{\mathrm{T}}{}_{j(y)})\\
&\quad + (E_{y1}H_{y2} - E_{y2}H_{y1})(-H_{y1}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)} + H_{x1}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(y)})]\\
&\quad /(H_{x1}H_{y2} - H_{x2}H_{y1})^2,\\
{}^{1}\mathbf{g}_j{}^{\mathrm{T}}{}_{yy} &= [(H_{x1}H_{y2} - H_{x2}H_{y1})(-E_{x2}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)} + H_{x2}{}^{\mathbf{e}}\mathbf{g}^{\mathrm{T}}{}_{j(x)})\\
&\quad + (E_{y2}H_{x1} - E_{y1}H_{x2})(-H_{x2}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(y)} + H_{y2}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)})]\\
&\quad /(H_{x1}H_{y2} - H_{x2}H_{y1})^2,\\
{}^{2}\mathbf{g}_j{}^{\mathrm{T}}{}_{yy} &= [(H_{x1}H_{y2} - H_{x2}H_{y1})(-H_{x1}{}^{\mathbf{e}}\mathbf{g}^{\mathrm{T}}{}_{j(y)} + E_{y1}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)})\\
&\quad + (E_{y2}H_{x1} - E_{y1}H_{x2})(-H_{y1}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(x)} + H_{x1}{}^{\mathbf{h}}\mathbf{g}^{\mathrm{T}}{}_{j(y)})]\\
&\quad /(H_{x1}H_{y2} - H_{x2}H_{y1})^2 .
\end{aligned} \tag{A10}$$

The eight vectors in eq. (A10) can be interpreted as generalized interpolator vectors, which involve linear combinations of vectors that interpolate the electric and magnetic fields from the forward modelling grid to the receiver at location j. These latter vectors are weighted by the combinations of electric and magnetic fields produced by the different source polarizations.

APPENDIX B: THE LINE SEARCH ALGORITHM

The line search procedure is to find α such that $\varphi(\mathbf{m} + \alpha\mathbf{u})$ is sufficiently decreased along the search direction, $\mathbf{u}$. Quadratic curve fitting is also employed in an attempt to refine α so that an even greater reduction in φ is possible. To determine α let us first normalize $\mathbf{u}$ so that

$$\mathbf{v} = \mathbf{u}/\|\mathbf{u}\|, \tag{B1}$$

where $\|\mathbf{u}\|$ is the Euclidean length, and define $\alpha' = \|\mathbf{u}\|\alpha$. The parameter α' has dimensions of $\mathrm{S\,m^{-1}}$, whereas α is dimensionless. We are now required to find α' such that $\varphi(\mathbf{m} + \alpha'\mathbf{v})$ is sufficiently decreased along the search direction, $\mathbf{v}$. Thus, the critical condition that α' must satisfy is

$$\varphi(\mathbf{m} + \alpha'\mathbf{v}) < \varphi(\mathbf{m}) + \delta\alpha'\nabla\varphi(\mathbf{m})\cdot\mathbf{v}, \tag{B2}$$

where δ is a small positive constant that ensures a sufficient decrease in $\varphi(\mathbf{m} + \alpha'\mathbf{v})$. The test in eq. (B2) is necessary since the simpler test,

$$\varphi(\mathbf{m} + \alpha'\mathbf{v}) < \varphi(\mathbf{m}), \tag{B3}$$

can lead to the possibility of oscillation about the solution without convergence (Dennis & Schnabel 1996). In choosing δ we follow recent optimization literature and set $\delta = 10^{-4}/\alpha'$. To launch the procedure we need to select a trial value for α' and carry out the test required by eq. (B2).

Unlike Newton and quasi-Newton methods, which accept a unit step ($\alpha' = 1$) most of the time, it is common to see step lengths that vary by one to two orders of magnitude with NLCG methods (Nocedal 1996). Therefore, we resort to a heuristic approach, which works well in practice for our problem. For the kth model parameter, m_k, and its perturbed value, $m_k + \Delta m_k$, for some point along the descent direction, $\mathbf{v}$, we have, from eq. (23),

$$m_k = m_0\,\mathrm{e}^{m'_k} + lb_k \tag{B4}$$

and

$$m_k + \Delta m_k = m_0\,\mathrm{e}^{m'_k + \Delta m'_k} + lb_k . \tag{B5}$$

Subtracting eq. (B5) from (B4) produces

$$\Delta m_k = m_0\,\mathrm{e}^{m'_k}(\mathrm{e}^{\Delta m'_k} - 1), \tag{B6}$$

and solving for $\Delta m'_k$ yields

$$\Delta m'_k = \ln[(\Delta m_k/m_0)\,\mathrm{e}^{-m'_k} + 1], \tag{B7}$$

where ln denotes the natural logarithm. By selecting the model parameter, $m_{\max}$, that corresponds to the largest component in $\mathbf{v}$ that satisfies the infinity norm,

$$\|\mathbf{v}\|_\infty = \max_{1 \le i \le M} |v_i|, \tag{B8}$$

we define the trial step based on eq. (B7) to be

$$\alpha'_{\text{trial}} = \ln[(1.6\mathrm{m}_{\max}/\mathrm{m}_0)\, \mathrm{e}^{-m'_{\max}} + 1], \tag{B9}$$

where $m_{\max}$ changes by a factor of 1.6. The factor 1.6 is empirical and is based upon the numerical experiments, which always demonstrated a sufficient decrease in the objective functional (eq. B2).

If eq. (B2) is satisfied with α'_{trial}, let $f_0 = \varphi(\mathbf{m})$ and $f_1 = \varphi(\mathbf{m} + \alpha'_{\text{trial}}\mathbf{v})$ and employ a quadratic model to find an α' that leads to an even greater reduction in f. Four pieces of information are required to define the quadratic: the two functional values $\varphi(\mathbf{m})$ and $\varphi(\mathbf{m} + \alpha'_{\text{trial}}\mathbf{v})$, α'_{trial} and the directional derivative $g_0 = \nabla\varphi(\mathbf{m}) \cdot \mathbf{v}$. Note that the directional derivative and functional at the point $\mathbf{m}$ are available from the prior NLCG iteration at no additional cost. Thus, if $f_{\min}$ defines the functional minimum we are seeking, then

$$f(x) = f_{\min} + b(x - \alpha')^2/{\alpha'_{\text{trail}}}^2, \tag{B10}$$

where

$$b = (f_1 - f_0) - g_0\alpha'_{\text{trial}}, \tag{B11}$$

then the candidate step is given by

$$\alpha' = -g_0\alpha'_{\text{trial}}/2b < \alpha'_{\max}. \tag{B12}$$

Here $\alpha'_{\max}$ sets an upper bound on the step, such that $\mathbf{m} + \alpha'\mathbf{v}$ leads to a realizable model and does not leave the domain of interest. For the step in eq. (B12) to be acceptable, two conditions must be met. The first is $b > 0$, which ensures that the quadratic model has positive curvature and α' defines its minimum and not its maximum, and the second that $f(\alpha')$ is actually less than f_1 by explicitly computing $\varphi(\mathbf{m} + \alpha'\mathbf{v})$. If $b < 0$ or if $f_1 < f(\alpha')$, we set $\alpha' = \alpha'_{\text{trial}}$ and exit the line search algorithm since we have already determined that α'_{trial} leads to a sufficient decrease in $\varphi(\mathbf{m} + \alpha'\mathbf{v})$.

If at the trial step f_1 fails to satisfy eq. (B2), a quadratic backtracking strategy is invoked from α'_{trial} until a sufficient decrease in f is observed. Backtracking is effective because $\mathbf{v}$ is a descent direction since $g_0 = \mathbf{v} \cdot \nabla\varphi < 0$ at $\varphi(\mathbf{m})$. The quadratic formula used to backtrack is given by

$$f(x) = f_0 + g_0 x + cx^2, \tag{B13}$$

where

$$c = (f_1 - f_0 - g_0\alpha'_{\text{trial}})/{\alpha'_{\text{trial}}}^2, \tag{B14}$$

and the candidate step is determined to be

$$\alpha' = \frac{g_0{\alpha'_{\text{trial}}}^2}{2(f_1 - f_0 - g_0\alpha'_{\text{trial}})}. \tag{B15}$$

Note that the curvature, c, in eq. (B14) is always positive, hence the quadratic model will always interpolate to a minimum. For α' to be acceptable, we must verify that eq. (B2) is satisfied. If not, we set $\alpha'_{\text{trial}} = \alpha'$ and continue backtracking until an acceptable step is found.

A danger of estimating α' using any type of polynomial approximation is that α' may be too near zero to be of much use in reducing φ and the inverse solution can stagnate as a result. Polynomial *safeguarding* prevents against this (*cf.* Dennis & Schnabel 1996). When $\alpha' < 0.1\alpha'_{\text{trial}}$ in any of the quadratic modelling procedures discussed above, we always set $\alpha' = 0.1\alpha'_{\text{trial}}$.